

Safe-Nation Vision: AI-Based Identity Protection and Consent-Driven Digital Content System

Mandadapu Srinadh, Peruri Rupesh, Katakota Hemanth Kumar, Chinagudaba Sai Chetan

Department of Artificial Intelligence & Machine Learning, Parul Institute of Engineering and Technology

Parul University, Vadodara, Gujarat, India

Under the Guidance of Prof. Viraj Patel, AI & ML Department

Abstract—Photographs and videos are shared online within minutes of being recorded, and a single clip often shows people who were never asked whether they agree to appear in it. Manual privacy controls put the burden on the affected person and usually act after the content is already public. This paper presents Safe-Nation Vision, an AI-based identity protection and consent-management system for digital images and videos. Uploaded media is analyzed, faces are located with computer-vision tools, and each face is converted to a facial embedding and compared with the embeddings of registered users. When a registered person is recognized, the system checks the privacy status of the content and sends a consent request. The person answers Allow, Allow Every Time or Don't Allow, and the media is published or restricted accordingly; unmatched faces are labelled Unknown. The design follows five layers, from user interface to notification and publishing control, and the project followed the Waterfall model. The paper describes the requirements, architecture, workflow and test plan of the system. The paper does not claim measured recognition accuracy or processing time. The contribution is a workflow in which publication depends on the approval of the people who appear in the media.

Index Terms—identity protection, consent management, face detection, face recognition, computer vision, digital privacy, content classification

I. INTRODUCTION

Photos and videos are now shared almost as soon as they are recorded. Social networks, video-sharing sites and messaging applications make publishing easy, and one clip often contains several people. Many of them never get the chance to say whether they are comfortable appearing in it.

A face is one of the most direct identifiers a person has. When a video showing someone's face is published without approval, the consequences can range from embarrassment to real misuse of that person's identity. Most platforms handle the problem after the fact, through reporting tools, tagging settings and takedown requests. These depend on the affected person finding the content first, and by then it may already have been viewed and copied. Automated moderation does not close the gap either. As noted in our project report, existing systems mostly look for harmful or inappropriate material, and they do not check whether the people shown have agreed to be published.

Computer vision and machine learning make an earlier check possible. Face detection can find every face in an image or video frame, and face recognition can compare each face with the stored representation of a registered person [1]. The same techniques involve biometric data, though, so they have to be used with care; the face-recognition literature stresses responsible and ethical deployment [1]. A system that uses recognition to protect privacy must therefore protect the biometric data it relies on.

This paper presents Safe-Nation Vision, an AI-based identity protection and consent-management system for digital images and videos. The idea is direct. Before media is published, the system finds the people in it, works out which of them are registered users, and asks those users for permission. Their answers then control whether the media is published or restricted. The contributions of this work are:

- a consent-driven workflow that links face recognition and identity matching to a publishing decision (Allow, Allow Every Time or Don't Allow);
- a layered architecture and module design covering upload, content analysis, face detection, face recognition, consent, notification, database and administration;
- a requirements specification with 18 functional requirements and measurable non-functional targets; and
- a test plan that runs from registration to administration, so that results can be added and checked against the stated requirements.

The paper is organized as follows. Section II states the problem and objectives, Section III reviews related work, and Sections IV to VI describe the proposed system, its architecture and methodology, and its implementation. Sections VII and VIII cover results and testing, and Sections IX to XII discuss the work, its limitations, future directions and conclusions.

II. PROBLEM STATEMENT AND OBJECTIVES

A. Problem Statement

The problem addressed here is the publication of digital media that shows identifiable people who have not been asked for permission. Three things make it hard. A single video may hold many people across thousands of frames, so checking by hand is not practical. The person affected is usually not the person who uploads, so the uploader has no direct way to know what the people in the video want. And any check that involves recognizing faces handles biometric data, which is sensitive by nature. The project therefore asks whether identity recognition can be used to route a consent request to each recognized person before the content goes public.

B. Objectives

The objectives of Safe-Nation Vision, taken from the project report, are:

1. to detect human faces in uploaded images and videos and ignore non-human objects;
2. to identify registered individuals by comparing facial embeddings with stored data, and to label unmatched faces as Unknown;

3. to check the privacy status of the content (public, private or moderate) and the permission status of each identified person;
4. to generate consent requests with a preview of the media and to collect the response;
5. to control publishing or visibility according to that response, blocking or blurring content when permission is denied;
6. to protect identity and biometric data through authentication, access control and encryption;
7. to keep secure records of uploads, consent responses and administrator overrides; and
8. to reduce manual monitoring by automating content analysis.

III. RELATED WORK

The literature survey in the project report covers twenty papers. Most of them come from intelligent surveillance and public safety, and only a few deal directly with faces or privacy. We group them by what they contribute to Safe-Nation Vision.

A. Video Surveillance and Anomaly Analysis

Most of the reviewed work applies computer vision to camera footage in order to detect events. Cucchiara and Grana [2] describe a multi-stage pipeline of adaptive background subtraction, object detection and multi-object tracking, followed by rule-based and statistical anomaly detection, with the aim of reducing the dependence of CCTV monitoring on human operators. Sultani et al. [3] address the shortage of labelled anomalies by combining multiple instance learning with deep CNN features and a ranking loss that produces anomaly scores. Datta and Roy [4] use CNNs with spatiotemporal features, optical-flow analysis and transfer learning to find violent or suspicious activity, and Hassner and Itcher [5] combine spatial and temporal learning in deep networks for violence detection. Ravanbakhsh and Nabi [6] apply CNNs, optical flow and crowd-density estimation to crowd behavior, and Ordóñez and Roggen [7] use deep networks to extract spatial and temporal features for human activity recognition. Olmos and Ferrer [8] train CNNs to detect firearms frame by frame.

The studies as summarized in the project report describe encouraging detection results, although the datasets and metrics differ and we do not compare them here. What these works share with Safe-Nation Vision is the way video is handled: frames are analyzed by CNN-based models. What they do not address is who appears in the footage or whether those people agreed to it. Their output is an event or anomaly label.

B. Face Detection and Recognition

Li and Jain [1] review face detection and recognition methods built on machine-learning models trained with large facial image datasets. They find recognition to be highly accurate under controlled conditions and stress that it should be deployed responsibly and ethically. Girshick and Taigman [9], as summarized in the project report, look at deep-learning face detection and recognition for surveillance and point out that footage often suffers from poor lighting, motion blur, low resolution and partial occlusion. They apply preprocessing such as normalization, alignment and noise filtering before recognition to make it more reliable.

These two works shape the face-related parts of our design. Recognition is the core identity mechanism, the image-quality step in the upload pipeline follows the idea of preprocessing, and the concern about responsible use carries over into how biometric data is handled.

C. Edge Computing, IoT and Privacy-Aware Processing

Ren and Zhang [10] propose processing video locally on edge devices and sending only metadata, alerts or summaries to the cloud, which lowers latency and bandwidth use and improves privacy. Mach and Becvar [11] also argue for processing data near its source, reporting lower latency and better reliability and bandwidth use than cloud-only models. Kumar and Patel [12] combine IoT cameras and sensors, edge preprocessing and cloud aggregation for smart-city surveillance, and Ahmed and Mahmood [13] use unsupervised learning to detect anomalies without labelled data. For our purposes [10] is the most relevant, because it treats limiting how much raw footage leaves a device as a legitimate privacy strategy. The current design of Safe-Nation Vision is server and cloud oriented, so this remains a direction to consider (Section XI).

D. Multimodal Sensing, Alerts and Platforms

A last group extends surveillance beyond video. Salamon and Bello [14] detect emergency sounds from audio features, and Chen and Sun [15] fuse video, audio and environmental sensor data. Imran and Castillo [16] mine social media text for early crisis signals, Puri and Verma [17] combine drones with AI video analytics, and Gupta and Banerjee [18] predict disasters from environmental sensor readings. Two papers deal with how detections reach people: Zhang and Wang [19] send automated alerts through SMS, control-room dashboards and emergency systems, and Lee and Park [20] describe a centralized national platform with a real-time dashboard. Most of this work lies outside consent management, but [19] and [20] are useful references for the notification and administrator-monitoring parts of Safe-Nation Vision.

TABLE I
RELATED WORK GROUPED BY AREA

Area	Refs.	Main contribution	Relevance to Safe-Nation Vision
Video surveillance and anomaly analysis	[2]–[8]	CNN and spatiotemporal analysis of video frames to detect anomalies, violence, crowd events, firearms and human activity	Supports frame-based CNN analysis of video; outputs are event labels, not identity or consent
Face detection and recognition	[1], [9]	Review of recognition methods; preprocessing for poor lighting, blur and occlusion; call for responsible use	Basis for detection, recognition, image-quality enhancement and careful handling of biometric data
Edge, IoT and privacy-aware processing	[10]–[13]	Local processing, less raw data sent to servers, lower latency, unsupervised anomaly detection	Privacy motivation; direction for later work on scalable deployment
Multimodal sensing, alerts and platforms	[14]–[20]	Audio, sensor fusion, social media, drones, disaster prediction, automated alerts, centralized dashboards	Informs notification and administrator monitoring; the rest is outside the current scope

E. Research Gap

Most of the surveyed studies use AI to detect events such as threats, anomalies or crises in footage that is already being captured. Where identity appears at all ([1], [9]), it is treated as a recognition task, and even there the discussion is about accuracy and responsible use rather than about the consent of the person recognized. The project report makes the same observation about existing systems: they concentrate on harmful or inappropriate content and lack a way to verify consent before an individual's image is published. In the works reviewed here, none obtains approval from a person who appears in media before that media goes public. Safe-Nation Vision is aimed at this gap. The survey is limited to the twenty papers in the report, and it does not cover work on consent management or privacy-preserving media publishing, which a broader review should include.

IV. PROPOSED SYSTEM

A. System Overview

Safe-Nation Vision acts as a middleware layer between the point where media is uploaded and the point where it is published. It takes an uploaded image or video, analyzes it, and decides whether the media can be published, must be restricted, or has to wait for approval. Three kinds of users interact with it: the uploader, the registered individual whose facial data is stored in the system, and the administrator (Fig. 2). The modules are listed in Table II and described below.

B. User Module

This module handles registration, login and authentication. Registered individuals provide facial data when they sign up, which is what makes later matching possible. Uploaders can see the consent status of their media and receive the final decision. The project report also describes role-based access levels (normal, moderated and restricted), email verification and face-recognition login.

C. Media Upload Module

The upload module accepts videos in MP4, AVI or MOV format, with a design limit of 1 GB, and images. For video it extracts frames at configurable intervals and passes them on for analysis.

D. Content Analysis Module

Before detection, the system checks image quality. Low-quality frames go through an enhancement step (the flowchart names Python with OpenCV), and good-quality frames skip it. The content is then classified as public, private or moderate

using AI models. As the design flowchart describes them, public content is safe content that can be shown at any time, private content is unsafe or age-restricted and is visible only if permission is granted, and moderate content is sensitive and permission-based. Faces in public content are still checked against stored allow and deny settings.

E. Face Detection Module

This module locates every visible human face in the extracted frames and ignores non-human objects. The design chapter names OpenCV and MediaPipe as the tools for this step. The SRS sets a detection accuracy target of at least 95%, which is a requirement and not a measured result.

F. Face Recognition Module

For each detected face the module generates a high-dimensional facial embedding, a numerical representation of the facial features. The embedding is compared with the stored embeddings of registered users using cosine similarity. A match is accepted only above a similarity threshold that the administrator can configure, and faces that do not match are labelled Unknown.

G. Identity Verification

A similarity score alone does not start the consent process. The matched identity is checked against the registered user's record in the database, and only then is a consent request issued for that person. The report does not describe any additional verification step, such as a manual check.

H. Consent Management Module

The consent module sends a request, with a preview of the media, to each identified user. The user can approve or reject. The design flowchart presents the options as Allow, Allow Every Time (or Allow Once for private content) and Deny. If consent is denied, publishing is blocked or the content is blurred, and the uploader is told about the decision. When several people are identified, the uploader has to wait for all of them, and publication depends on their combined responses.

I. Notification and Permission Module

Consent requests, status updates and decisions reach users through notifications. The design lists WebSocket communication (Socket.io) for live updates, and the SRS allows email or SMS services. The same module applies stored permission choices, so that a person does not have to answer identical requests repeatedly.

J. Database Module

The database holds user details, facial embeddings, uploaded media, classification results, consent records and logs. The SRS asks for indexed storage that supports high-dimensional vectors so that embeddings can be retrieved quickly, automated backups every 24 hours, and retention policies that follow privacy regulations.

K. Administration Module

Administrators have a monitoring panel showing detection results, consent status, flagged content and system logs. They manage the database, configure the similarity threshold and handle disputes. An administrator can override a decision, and every override is recorded with a timestamp.

TABLE II
SYSTEM MODULES

Module	Main function
User	Registration, login, authentication; facial data registration
Media upload	Accepts images and MP4/AVI/MOV video; extracts frames
Content analysis	Quality check and enhancement; public, private or moderate classification
Face detection	Finds all visible human faces in frames
Face recognition	Facial embeddings, cosine-similarity matching, Unknown label
Identity verification	Confirms the match against the database record
Consent management	Sends requests with preview; records responses; blocks or restricts on denial
Notification and permission	Delivers requests and updates; applies stored permissions
Database	Users, embeddings, media, classification results, consent records, logs
Administration	Monitoring, threshold configuration, disputes, logged overrides

TABLE III
FUNCTIONAL REQUIREMENTS FROM THE SRS

Group	IDs	Requirement summary
Video processing	SV-REQ-1 to 3	Accept MP4, AVI, MOV; extract frames at configurable intervals; videos up to 1 GB
Face detection	SV-REQ-4 to 6	Detect all visible faces; ignore non-human objects; accuracy target of at least 95%
Face recognition	SV-REQ-7 to 10	Facial embeddings; cosine similarity; Unknown label; admin-configurable threshold
Consent management	SV-REQ-11 to 15	Send requests with preview; approve or reject; block publishing on denial; notify uploader
Audit and logging	SV-REQ-16 to 18	Log uploads and consent responses; record admin overrides with timestamp

V. SYSTEM ARCHITECTURE AND METHODOLOGY

A. Layered Architecture

The system is organized into five layers, shown in Fig. 1. The user interface layer is the web or mobile front end, where users register, log in, upload media and respond to consent requests, and where administrators monitor activity. Requests travel over HTTPS to the application logic layer, which

handles authentication (JWT or OAuth) and coordinates the modules through REST APIs. Media that needs analysis goes to the AI processing layer, which performs frame extraction, quality enhancement, face detection, embedding generation and matching. Matching needs stored embeddings, so this layer reads from the database layer, which also holds user records, media, classification results and consent records. The result returns to the application layer, which starts the consent workflow. The last layer, notification and publishing control, delivers requests and updates and applies the final decision to publish, restrict or block. The report describes the design as modular and client-server based, with cloud services for scalable storage and processing.

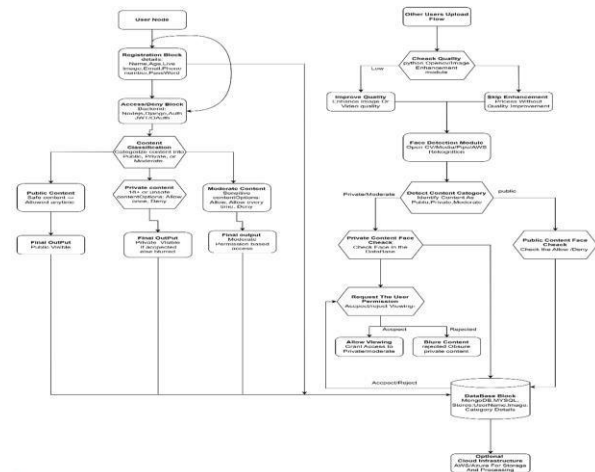


Fig. 1. System architecture of Safe-Nation Vision.

B. Processing Workflow

The workflow that every uploaded item follows is summarized below:

- Media upload.** The uploader submits an image or video, and frames are extracted from video.
- Content analysis.** Image quality is checked, enhanced if necessary, and the content is classified.
- Face detection.** All visible faces are located in the frames.
- Face recognition and identity matching.** Each face is converted to an embedding and compared with registered users.
- Database verification.** A matched identity is confirmed against the stored user record; unmatched faces are labelled Unknown.
- Privacy and permission check.** The system looks at the content category and the permission status of each identified person.
- Consent request.** A request with a preview is sent to each identified person.
- Response.** The person chooses Allow, Allow Every Time or Don't Allow.
- Controlled publishing or restriction.** The media is published if permission is given, and blocked or blurred if it is not.

Two points are not fully defined in the report. The report does not define the scope of Allow Every Time, for example whether it applies to one uploader or to all, or how faces

labelled Unknown are handled and whether they block publication.

C. Use Cases and Data Flow

Fig. 2 shows the three actors and their use cases. The uploader registers, logs in and uploads media, the registered individual receives the consent request and approves or rejects it, and the administrator monitors the system. Inside the system, an upload passes through frame extraction, face detection, face recognition and content classification before a request is issued, and the outcome ends in publishing or blocking, with the data stored.

The data flow diagrams describe the system at three levels. At Level 0 (Fig. 3) it is a single process exchanging data with the user, the administrator and the database. Level 1 (Fig. 4) splits it into login and registration, upload, face detection, face recognition and consent management, all working with the system database. A Level 2 diagram would break the consent-management process into its steps, but the report does not include one.

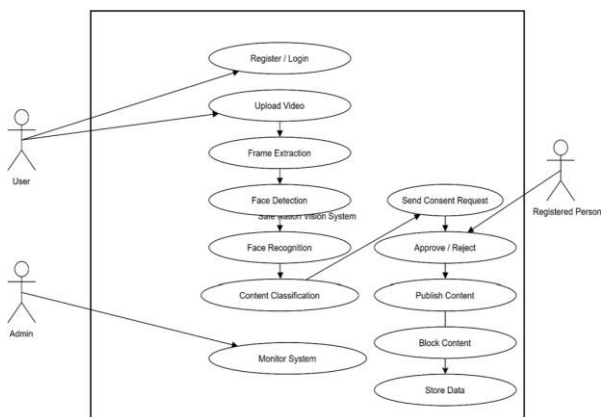


Fig. 2. Use case diagram of Safe-Nation Vision.



Fig. 3. Data flow diagram, Level 0 (context diagram).

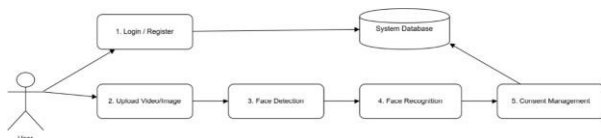
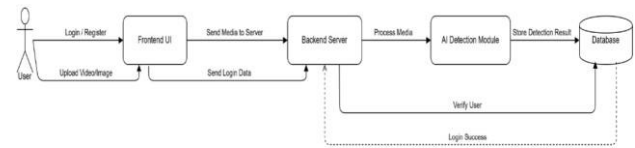


Fig. 4. Data flow diagram, Level 1.

D. Interaction and Activity Flow

The sequence in Fig. 5 follows a request through the front end, backend server, AI detection module and database, and returns the consent outcome to the uploader. Fig. 6 shows the same flow as activities and decisions, including the branch for faces that are not recognized.



F6.Sequence Diagram

Fig. 5. Sequence diagram of the upload and consent process.

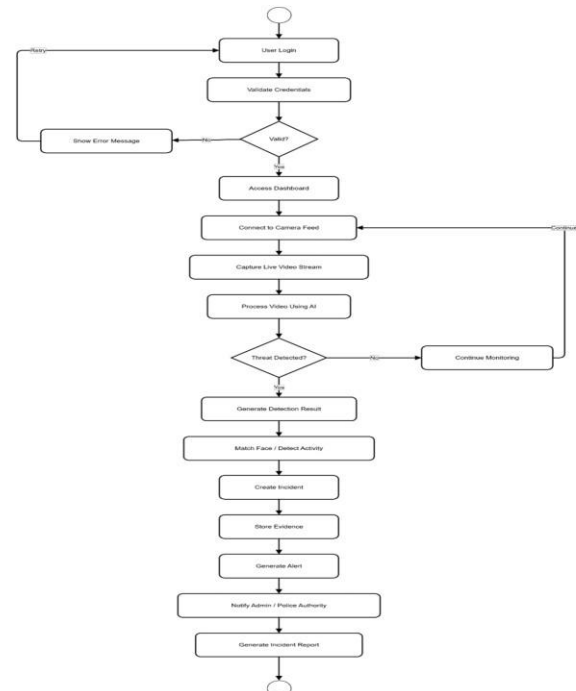


Fig. 6. Activity diagram of the consent workflow.

E. Development Methodology

The project followed the Waterfall model, in which each phase is finished before the next begins. The report gives the reason as follows: the requirements were clear at the start, the main functional modules were fixed, and the model suits a student project that needs proper documentation and easy progress tracking. Agile, Spiral, Prototype and V-Model were considered less suitable because they pay off when requirements change often or risk analysis is critical. The six phases were:

1. **Requirement analysis.** Face detection, identity recognition, consent verification and secure publishing were identified as the needs, along with technical needs such as languages and database.
2. **System design.** The layered architecture and the workflow from upload to publish or block were defined and documented with UML diagrams and flowcharts.
3. **Implementation.** The modules were coded and integrated.
4. **Testing.** Unit, integration, system, performance and security testing were planned.
5. **Deployment.** The report describes deployment on a cloud platform (AWS, Azure or Google Cloud). The report does not say whether the prototype has been deployed.
6. **Maintenance.** Monitoring, bug fixes, model updates and security patches.

Before the design phase, the team studied face detection and recognition approaches, including CNNs, FaceNet and Vision Transformers, and looked at the privacy and ethical issues of publishing media. The report names FaceNet and Vision Transformers without citing them. It does not state that the team trained any of these models.

IMPLEMENTATION

The report describes a web application with an AI processing back end. Several layers are listed with alternative technologies, and Table IV shows them side by side. We do not treat all of the options as used. The report does not state which option was finally used for each layer, so the table lists the options without claiming that all of them were adopted.

TABLE IV
TECHNOLOGY STACK NAMED IN THE PROJECT REPORT

Layer	Technologies named in the report
Frontend	HTML, CSS, JavaScript; React.js, Angular or Vue.js
Backend	Python (Flask or Django) or Node.js with Express.js
API	RESTful APIs over HTTPS (Express.js or Django REST Framework)
Image processing	OpenCV (quality enhancement and preprocessing)
Face detection	MediaPipe with OpenCV; AWS Rekognition also appears in the flowchart
Recognition and classification models	TensorFlow or PyTorch; deep-learning embeddings compared by cosine similarity; TensorFlow/AI models for content classification
Database	MySQL or MongoDB; encrypted, indexed, with support for high-dimensional vectors
Notification	WebSocket with Socket.io; email or SMS services
Authentication	JWT, OAuth; multi-factor authentication for administrators
Cloud	AWS, Microsoft Azure or Google Cloud Platform

A. Frontend

The front end is a web interface for desktop and mobile browsers. Its main screens are login and registration, a secure upload page, a consent dashboard for identified users and an administrator monitoring panel. The report's abstract mentions JavaScript and Visual Studio as development tools.

B. Backend and APIs

The backend receives uploads, coordinates the AI modules and manages consent state. The report's implementation chapter refers to Python for the backend, while the design chapter lists Python and Node.js as options. Communication between the front end, backend and AI modules uses REST APIs over HTTPS.

C. Database and Storage

The database stores user records, facial embeddings, media, classification results, consent records and audit logs. The SRS requires AES-256 encryption for biometric data, indexed retrieval of embeddings, and automated backups. The entity-relationship diagram in the project report describes cameras, incidents, alerts and police authorities, which do not belong to the consent workflow, so the entities here are limited to those listed above.

D. Image and Video Processing

Uploaded videos are converted to frames at intervals the system can configure. Frames and images pass a quality check, and low-quality ones are enhanced before detection. The design chapter names OpenCV for this work, in line with the preprocessing idea in [9].

E. Face Detection and Recognition

The design chapter names MediaPipe for face detection. Recognition relies on deep-learning embeddings compared by cosine similarity against a configurable threshold, and the SRS names TensorFlow or PyTorch as the AI framework. The report does not say which embedding model the prototype uses, or whether the team trained a model or used a

pretrained one, and no training or evaluation dataset is described.

F. Authentication and Security

Authentication is planned with JWT or OAuth, with role-based access control and multi-factor authentication for administrators. Communication is meant to be encrypted, and consent records must be stored securely. The SRS lists these as requirements. The report does not say which of them the prototype implements.

G. Consent, Notification and Deployment

Consent requests and outcomes are delivered as notifications, with Socket.io listed for live updates and email or SMS as optional channels. For deployment the design assumes a cloud platform with GPU-enabled servers, load balancing and horizontal scaling. The report does not describe the final deployment.

VI. RESULTS AND ANALYSIS

At the time of writing, the project report does not contain measured results. It gives no values for detection accuracy, recognition accuracy, processing time or dataset size, and none are reported here. This section therefore separates what the report does establish from what still has to be measured.

A. Design-Level Outcomes

The report establishes a complete workflow from upload to controlled publishing, a five-layer architecture, a set of modules with 18 identified functional requirements (SV-REQ-1 to SV-REQ-18), and non-functional targets for performance, security, reliability, usability, scalability and maintainability. These are design outcomes. They show what the system is meant to do, not how well it does it.

B. Performance Targets

The SRS defines numerical targets rather than measured outcomes. It calls for face detection accuracy of at least 95 percent, identity recognition accuracy of at least 93 percent,

a response time per face of no more than three seconds, frame processing latency of no more than five seconds, support for at least 100 concurrent uploads, and system uptime of at least 99 percent. These figures describe what the system is meant to achieve. No measurements against any of them have been carried out, so none of these numbers should be read as an achieved result.

C. Functional Observations

Several workflow stages are meant to be observed once the system is exercised: whether a supported file is accepted and its frames extracted, whether every visible face in a frame is found, whether a registered person is correctly matched while others are left labelled Unknown, whether a consent request with a media preview reaches the identified person, whether each consent option is recorded correctly, whether publishing follows consent and blocks or blurs content on denial, and whether uploads and responses are logged. None of these stages has been observed and recorded yet, so this section describes what the system is designed to do rather than what it has been shown to do.

D. Analysis

Until these stages have actually been observed, the results described above should be read as a design result, not a performance claim. When measurements are eventually taken, we suggest reporting detection accuracy and recognition accuracy separately, because errors at the two stages have different consequences. A face that is missed produces no consent request. A wrong match sends the request to the wrong person. The literature also suggests testing under different conditions, since recognition is reported to work well in controlled settings [1] and to suffer from poor lighting, blur and occlusion [9].

VII. TESTING AND VALIDATION

The report plans five kinds of testing: unit, integration, system, performance and security. No test has been executed against the prototype, and the project report contains no test evidence, so the subsections below describe the intended scope of each kind of testing rather than results.

A. Unit Testing

Unit testing is intended to check individual modules in isolation, including registration and login, media upload and frame extraction, face detection, and face recognition and identity matching. For registration, the check would be that an account is created and facial data is stored as an embedding. For login, valid credentials should grant access and invalid credentials should be rejected. For media upload, MP4, AVI, MOV and image files should be accepted and frames extracted from video. For face detection, all visible faces in a frame should be found while non-human objects are ignored. For face recognition, a registered person should be matched above the configured similarity threshold, and an unmatched face should be labelled Unknown.

B. Integration Testing

Integration testing is intended to confirm that the modules work together correctly. Examples include checking that a consent request carrying a media preview reaches the person identified by the recognition module, that each consent response, Allow, Allow Every Time or Don't Allow, is recorded correctly and applied to publishing, and that a

request involving several identified people waits for all of their responses before the media is published.

C. System Testing

System testing is intended to check the complete system against the requirements listed in the SRS, following the workflow from upload through to publishing or restriction. This includes verifying that media with a pending or denied consent decision cannot be published or is blurred instead, and that database operations correctly store and retrieve facial embeddings and consent records.

D. Security Testing

Security testing is intended to check that unauthenticated or unauthorized requests are rejected, that administrator functions such as monitoring, threshold configuration and overrides work as intended, and that every administrator override is logged with a timestamp so the process remains auditable. It would also cover audit logging of uploads and consent responses more generally.

E. Testing Limitations

None of the tests described above has been executed, and the project report contains no pass or fail evidence for any of them. The prototype has therefore not been validated against its own requirements, and the testing described here should be read as a plan for validation rather than a record of it.

VIII. DISCUSSION

The main design decision in Safe-Nation Vision is to move the privacy check from after publication to before it. Manual controls such as tagging settings and takedown requests depend on the affected person noticing the content. Here the system finds the person and asks. Automated detection is what makes this workable for video, since a clip may hold many faces across many frames and checking them by hand at upload time would not be practical.

Recognition and consent are tightly coupled, and that coupling decides where errors matter. A missed face means no request is sent and the person is exposed without being asked. A false match sends the request to the wrong person, whose answer then stands in for someone else's. This is why the SRS makes the similarity threshold configurable by the administrator and labels unmatched faces as Unknown instead of forcing a match. The right threshold can only be chosen from measured error rates, and those are not yet available. The literature adds a caution: recognition is accurate in controlled conditions [1], and preprocessing is needed for poor lighting, blur and occlusion [9].

Because the system stores facial embeddings, it creates a privacy risk of its own. The SRS therefore asks for AES-256 encryption of biometric data, encrypted communication, role-based access, multi-factor authentication for administrators, secure storage of consent records, and no sharing of biometric data with third parties without authorization. These are requirements, and the report does not say how many of them the prototype implements. Local processing of the kind proposed in [10] could reduce exposure further.

Responsibility is shared. The AI performs detection, recognition and content analysis, and people make the consent decisions. Administrators can override a decision, but every override is logged with a timestamp, which keeps the

process auditable. The system can only protect people who have registered. Someone who has never registered is labelled Unknown, and the system has no way to ask them. This is a design boundary rather than a defect, but anyone deploying the system should be told about it.

IX. LIMITATIONS

The limitations below follow from the project report and the literature it uses.

- No quantitative evaluation has been reported yet, so the SRS performance targets described in Section VII are unverified.
- Recognition accuracy depends on image and video quality, lighting, pose and occlusion [1], [9]. Poor frames can lead to missed faces or wrong matches.
- Only registered individuals can be identified. Unregistered people remain Unknown and cannot be asked for consent.
- Processing large videos, with a design limit of 1 GB, is computationally heavy. GPU-enabled servers are expected, and the latency targets have not been verified.
- Encryption, backup, load balancing and scaling are specified in the SRS. The report does not confirm production-level hardening of the prototype.
- Performance depends on the chosen detection and recognition models or services, and the report does not state which were finally used.
- Some behavior is still undefined, including the scope of Allow Every Time, the handling of Unknown faces and what happens when a person does not reply.
- Consent delivery assumes stable internet connectivity and working notification services.

X. FUTURE WORK

The following directions come from the project report and are **future work**. None of them is implemented in the current system.

- **Real-time video consent validation:** checking consent on streaming video.
- **Deepfake detection:** identifying manipulated media.
- **Improved face-recognition models:** more advanced deep-learning models for better accuracy.
- **Privacy-preserving processing:** for example local processing, as in [10], so that raw media leaves the user's device less often.
- **Blockchain-based consent verification:** tamper-evident consent records.
- **Voice recognition:** identity checks that use voice as well as faces.
- **Scalable cloud deployment:** with real-time mobile alerts and cloud-based monitoring.

The report also lists surveillance-oriented extensions, such as activity recognition, weapon detection, audio anomaly detection, social-media crisis detection and drone integration. They would take the system beyond identity protection, so we leave them outside the scope of this paper.

XI. CONCLUSION

This paper described Safe-Nation Vision, an AI-based system that ties identity recognition to consent for digital images and videos. Faces are detected, converted to embeddings and matched against registered users. Those users are asked to allow or deny publication, and their answers decide whether the media is published or restricted. This connection between recognition and consent is what distinguishes the system from surveillance and moderation tools that only detect events or harmful content.

What the project has established so far is the workflow, the layered architecture, the module and requirement specification, and a test plan. Measured accuracy, latency and test results are not yet available, so the performance targets remain targets. The system does not solve identity misuse as a whole. It protects only registered individuals, depends on recognition quality, and needs production-level security before real use. Future work includes real-time consent validation, deepfake detection, stronger recognition models and privacy-preserving processing.

ACKNOWLEDGMENT

The authors thank Prof. Viraj Patel for guidance throughout the project and Dr. Kamal Sutaraiya, Head of the AI&ML Department, for support and feedback.

REFERENCES

- [1] S. Z. Li and A. K. Jain, "Facial recognition systems for law enforcement applications," *IEEE Computer*, vol. 44, no. 1, pp. 40–47, 2022.
- [2] R. Cucchiara and C. Grana, "Intelligent video surveillance for public safety," *IEEE Trans. Syst., Man, Cybern.*, vol. 40, no. 6, pp. 1721–1733, 2010.
- [3] Y. Sultani, C. Chen, and M. Shah, "Real-time anomaly detection in surveillance videos," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2018, pp. 6479–6488.
- [4] A. Datta and K. Roy, "Deep learning for crime detection in surveillance videos," *Pattern Recognit. Lett.*, vol. 105, pp. 256–263, 2018.
- [5] T. Hassner and Y. Itcher, "Violence detection in surveillance videos using deep neural networks," *Pattern Recognit.*, vol. 94, pp. 321–331, 2019.
- [6] S. Ravanbakhsh and M. Nabi, "Crowd behavior analysis for public safety using deep learning," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 11, pp. 3564–3575, 2018.
- [7] F. Ordóñez and D. Roggen, "Human activity recognition using deep learning for smart surveillance," *Sensors*, vol. 16, no. 1, pp. 1–22, 2016.
- [8] A. Olmos and J. C. Ferrer, "Gun detection in surveillance cameras using deep learning," *Appl. Sci.*, vol. 8, no. 9, pp. 1–15, 2018.
- [9] R. Girshick and Y. Taigman, "Deep learning-based face detection and recognition in intelligent surveillance systems," *Int. J. Comput. Vis. Pattern Recognit.*, vol. 18, no. 4, pp. 114–125, 2021.
- [10] J. Ren and H. Zhang, "Privacy-preserving smart surveillance using edge AI," *IEEE Internet Things J.*, vol. 17, no. 9, pp. 8448–8458, 2020.
- [11] P. Mach and Z. Becvar, "Edge computing for intelligent surveillance systems," *IEEE Commun. Mag.*, vol. 55, no. 8, pp. 102–108, 2017.
- [12] S. Kumar and P. Patel, "Smart city surveillance using IoT and artificial intelligence," *Int. J. Adv. Comput. Sci. Appl.*, vol. 11, no. 4, pp. 215–223, 2020.
- [13] M. Ahmed and A. Mahmood, "Anomaly detection in smart cities using machine learning," *J. Big Data*, vol. 6, no. 1, pp. 1–24, 2019.
- [14] J. Salamon and J. P. Bello, "Audio-based anomaly detection for public safety systems," *IEEE Signal Process. Lett.*, vol. 24, no. 8, pp. 1200–1204, 2017.
- [15] K. Chen and L. Sun, "Multimodal sensor fusion for intelligent public safety systems," *Sensors*, vol. 20, no. 6, pp. 1742–1754, 2020.
- [16] A. Imran and C. Castillo, "Social media mining for real-time crisis detection," *ACM Trans. Inf. Syst.*, vol. 35, no. 1, pp. 1–28, 2017.

- [17] K. Puri and R. Verma, "Drone-based surveillance systems for public safety," *J. Intell. Robot. Syst.*, vol. 95, no. 3, pp. 567–580, 2019.
- [18] A. Gupta and R. Banerjee, "Disaster detection and response using artificial intelligence," *Int. J. Disaster Risk Reduct.*, vol. 45, Art. no. 101112, 2020.
- [19] L. Zhang and H. Wang, "AI-based real-time alert system for public safety," *Future Gener. Comput. Syst.*, vol. 4, no. 6, pp. 463–472, 2019.
- [20] D. Lee and S. Park, "AI-driven national public safety platforms," *Gov. Inf. Q.*, vol. 37, no. 2, pp. 101–109, 2020.